⑲日本国特許庁(JP) ⑪特許出願公開

⑫公開特許公報(A) 平3-288934

⑤Int. Cl.[5]	識別記号	庁内整理番号	㊸公開 平成3年(1991)12月19日
G 06 F 9/46	350	8120-5B	

審査請求 未請求 請求項の数 1 (全7頁)

㊹発明の名称 仮想計算機システムにおけるデータ転送制御方式

㉑特 願 平2-90590
㉒出 願 平2(1990)4月5日

㉒発 明 者 斎 藤 優 神奈川県川崎市中原区上小田中1015番地 富士通株式会社内
㉑出 願 人 富士通株式会社 神奈川県川崎市中原区上小田中1015番地
㉔代 理 人 弁理士 小笠原 吉義 外2名

明 細 書

1. 発明の名称

仮想計算機システムにおける
データ転送制御方式

2. 特許請求の範囲

CPUの非同期転送命令によってアクセス可能な外部記憶装置(18)を備えた仮想計算機システムにおけるデータ転送制御方式において,

各仮想計算機(10)の転送優先度に関する定義情報を記憶する手段(14)と,

各仮想計算機上で動作するオペレーティング・システム(11)からの主記憶装置と外部記憶装置間のデータ転送要求をキューイングする手段(13)と,

各仮想計算機の転送優先度に従って,1回の転送データ長を短く制限する手段(15)と,

短く制限した転送データ長でオペレーティング・システムからのデータ転送要求を代行する手段(16)とを備え,

オペレーティング・システムからの要求が満足されなかった場合には,仮想計算機モニタ(12)内で自動的にまたはオペレーティング・システムからの要求により,再度不足分の要求をキューイングし,データを分割して転送するようにしたことを特徴とする仮想計算機システムにおけるデータ転送制御方式。

3. 発明の詳細な説明

〔概要〕

各仮想計算機上で動作するOSの外部記憶装置に対するアクセス性能を,仮想計算機モニタにより制御できるようにした仮想計算機システムにおけるデータ転送制御方式に関し,

外部記憶装置に対する大量のデータ転送により,他の仮想計算機の要求が長時間待たされることがないようにすることを目的とし,

各仮想計算機の転送優先度に関する定義情報を記憶する手段と,オペレーティング・システムからの主記憶装置と外部記憶装置間のデータ転送要求をキューイングする手段と,転送優先度に従っ

て，１回の転送データ長を短く制限する手段と，短く制限した転送データ長でデータ転送要求を代行する手段とを備え，オペレーティング・システムからの要求が満足されなかった場合には，仮想計算機モニタ内で自動的にまたはオペレーティング・システムからの要求により，再度不足分の要求をキューイングし，データを分割して転送するように構成する。

〔産業上の利用分野〕

本発明は，各仮想計算機上で動作するオペレーティング・システム（ＯＳ）の外部記憶装置に対するアクセス性能を，仮想計算機モニタにより制御できるようにした仮想計算機システムにおけるデータ転送制御方式に関する。

コンピュータシステムの大規模化および高速化の要求に伴い，外部記憶装置によるシステム内／システム間の高速・大量データ転送が実現されようとしている。仮想計算機システムにおいても同じ要求があり，各仮想計算機（ＶＭ）ごとに，外部記憶装置を割り当てることにより，外部記憶装置による仮想計算機内／仮想計算機間のデータ転送を実現しようとしている。

しかし，データ転送量が多い場合には，データ転送処理が，ある特定の仮想計算機に偏ってしまい，性能上の問題が発生するおそれがある。

そのため，仮想計算機モニタにより，各仮想計算機からの要求を，バランスよく制御できるデータ転送制御方式が必要となる。

〔従来の技術〕

仮想計算機システムではないネイティブ環境のシステムにおいては，外部記憶装置と主記憶装置との間のパス数は，一般にシステムの数より少なくなることはない。そのため，主記憶装置と外部記憶装置間の転送に用いるパスが，他システムで使用されていることにより，転送要求が待たされることはない。

ところが，仮想計算機システムでは，主記憶装置と外部記憶装置間のパス数が，仮想計算機の数よりも少ないことがあり得るため，ある仮想計算機が大量のデータを外部記憶装置へ転送すると，その仮想計算機が主記憶装置と外部記憶装置間のパスを長時間専有することになり，専有できなかった他の仮想計算機の転送要求が大幅に遅れ，転送が完了するまで長時間待たされることが発生する。

〔発明が解決しようとする課題〕

したがって，仮想計算機上で動作するＯＳが，例えばページング処理などのＯＳ全体の性能に影響する箇所で，主記憶装置と外部記憶装置間の転送を行っている場合などは，ＯＳの性能に大きな影響を与えてしまうことになり，結果として仮想計算機ごとの性能に大きな偏りができてしまう。

本発明は，各仮想計算機上で動作するＯＳの外部記憶装置に対するデータ転送の性能を，仮想計算機モニタにより制御できるようにし，外部記憶装置に対する大量のデータ転送により，他の仮想計算機の要求が長時間待たされることがないようにすることを目的としている。

〔課題を解決するための手段〕

第１図は本発明の原理説明図である。

第１図において，１０は仮想計算機（ＶＭ），１１は各仮想計算機１０上で動作するオペレーティング・システム（ＯＳ），１２は仮想計算機１０を制御するＶＭモニタ，１３は転送要求キューイング部，１４は転送優先度記憶部，１５は転送長制限部，１６は転送命令代行部，１７は転送残り長算出部，１８は外部記憶装置，１９は主記憶装置，２０はパスを表す。

外部記憶装置１８は，例えば半導体で構成され，高速にアクセスすることができる記憶装置であって，主記憶装置１９とは別に設けられるものである。外部記憶装置１８に対するアクセスは，入出力命令によらずに，ＣＰＵの非同期転送命令によって行うことができるようになっている。

パス２０は，外部記憶装置１８と主記憶装置１９間のデータ転送経路であって，それぞれが並行

にデータを転送できるようになっているものである。例えばパス２０が３本あれば，同時に最大で３つのデータ転送を遂行できる。

本発明では，あらかじめ各仮想計算機１０の構成定義情報をシステムに登録する際に，各仮想計算機１０の転送優先度を登録できるようになっている。転送優先度記憶部１４は，その各仮想計算機１０ごとの転送優先度を記憶する。

仮想計算機１０で動作するオペレーティング・システム１１が，外部記憶装置１８と主記憶装置１９間の非同期転送命令を発行すると，ＶＭモニタ１２がそれをインタセプトし，転送要求キューイング部１３を起動する。

転送要求キューイング部１３は，その非同期転送命令に関する要求を内部のキューにつなぎ込む処理を行う。

実際のデータ転送処理は，キューから要求を１つずつ取り出して行う。取り出した要求について，転送長制限部１５は，転送優先度記憶部１４に記憶されている要求元の仮想計算機１０の転送優先度に従って，１回の転送データ長を短く制限する。すなわち，要求された転送データ長が，転送優先度に応じた転送データ長の制限よりも長い場合，要求された転送データ長を分割し，データ転送長の制限以下になるように短くする。

なお，転送優先度記憶部１４には，各仮想計算機１０ごとに制限する転送データ長の数値を直接的に記憶するようにしてもよい。

転送命令代行部１６は，短く制限した転送データ長で，仮想計算機１０のオペレーティング・システム１１の代わりに，非同期転送命令を発行する。これにより，主記憶装置１９と外部記憶装置１８間のデータ転送が開始される。

このデータ転送が完了したならば，転送残り長算出部１７により，要求された転送データ長のうち，まだ転送していない転送データ長を計算し，オペレーティング・システム１１からの要求が満足されたかどうかを判定する。

オペレーティング・システム１１からの要求が満足されなかった場合には，ＶＭモニタ１２内で，再度不足分の転送データ長の転送要求を，転送要求キューイング部１３が管理するキューに，再キューイングする。

または不足分の転送データ長を要求元のオペレーティング・システム１１に通知する。この場合，オペレーティング・システム１１は，通知された情報に基づき，非同期転送命令により，転送されなかった分についてのデータ転送を再度要求する。

オペレーティング・システム１１からの要求が満足された場合には，オペレーティング・システム１１に対して，転送の完了を通知する。

〔作用〕

本発明では，主記憶装置１９と外部記憶装置１８間のパス２０を使用する転送要求の転送データ長を，各仮想計算機１０に対して定義された転送優先度に見合った転送データ長で分割し，途中で他の仮想計算機１０の転送要求を割り込ませることができるように，ＶＭモニタ１２により，各仮想計算機１０のデータ転送処理を制御する。

これにより，ある１つの仮想計算機１０が主記憶装置１９と外部記憶装置１８間のパスを長時間専有することが防止される。したがって，仮想計算機１０上で動作するオペレーティング・システム１１が，オペレーティング・システム全体の性能に影響する箇所で，主記憶装置１９と外部記憶装置１８間のデータ転送を行っているときでも，仮想計算機１０ごとの転送に関係する性能に大きな偏りができることがなくなる。

〔実施例〕

第２図は本発明の第一実施例処理構成図，第３図は本発明の第一実施例処理フロー，第４図は本発明の第二実施例処理構成図，第５図は本発明の第二実施例処理フロー，第６図は従来技術と本発明を比較するためのタイムチャートを示す。

第２図および第３図に従って，本発明の第一実施例を説明する。以下の説明における①～⑪は，第２図および第３図に示す①～⑪に対応する。

① 仮想計算機ＶＭ－Ａで動作するＯＳが，非同

期転送命令を発行し，主記憶装置１９と外部記憶装置１８間のデータ転送を要求すると，その非同期転送命令をインタセプトし，転送要求キューイング部１３により，その要求をキューイングする。この要求を表す要求データブロック２１には，転送元アドレス，転送先アドレス，転送長等の情報が格納される。

② キューイングされた要求は，逐次，非同期転送制御部２３により取り出される。

③ その要求について，パス選択部２２は，主記憶装置１９と外部記憶装置１８間の空きパスを選択する。すなわち，外部記憶装置１８のアドレスによって，複数のパス２０が使用可能になっている場合に，各パス２０について空いているか否かを判定し，空いているパス２０を割り当てる。

④ 次に，転送長制限部１５により，転送優先度記憶部１４を参照し，各仮想計算機ごとの転送優先度に応じた１回当たりの転送長を算出する。

⑤ 非同期転送命令を，要求元のＯＳの代わりに発行し，データ転送を開始する。このデータ転送中は，ＯＳからの他の転送要求を，転送要求キューイング部１３によって受け付けることが可能である。

⑥ 仮想計算機ＶＭ－ＡのＯＳが要求する転送が完了するまでの間，例えば，仮想計算機ＶＭ－Ｂで動作するＯＳから転送要求があれば，転送要求キューイング部１３により，その要求をキューイングする。

⑦ 転送完了時にその割込み通知を受ける。

⑧ 転送が完了すると，要求された転送データ長とＶＭモニタにより代行した転送データ長を比較し，転送残り長ｚを算出する。

⑨ この転送残り長ｚが０かどうかにより，ＯＳからの転送要求長が満足したかどうかを判定する。

⑩ 判定の結果，ＯＳから要求されたデータ転送がすべて終了していれば，要求元のＯＳに転送の完了を通知する。

⑪ 転送残り長ｚが０でない場合には，その残りのデータ長ｚの要求を作成し，転送要求キューイング部１３が管理するキューの最後に，再度キューイングして，ＯＳからの要求が満たされるまで，ＶＭモニタ内で以上の処理を繰り返す。

第４図および第５図は，本発明の第二実施例を示している。

前述した第一実施例では，ＯＳからの要求が１回の転送で満足されなかった場合に，ＶＭモニタ内で自動的に要求を再キューイングし，すべての転送が完了してから，ＯＳに転送の完了を通知する。

これに対し，この第二実施例では，ＯＳからの要求が１回の転送で満足されなかった場合に，ＶＭモニタにより，ハードウェア仕様またはＯＳとのハンドシェイクを使用して，ＯＳからの要求が満足されなかったこと，および転送されなかった残りのデータ情報をＯＳに通知する。

ＯＳは，その情報を基に，転送されなかったデータについて，再度データ転送要求を行う。

最終的にすべてのデータ転送が終わると，ＯＳに完了が通知される。

第４図および第５図に示す処理では，①～⑧が第２図および第３図に示す①～⑧と同様である。第二実施例の場合，処理⑨では，転送残り長ｚの情報を作成し，転送残り長ｚが０か否かにより，ＯＳに対して完了通知または不足情報の通知を行う。ＯＳは，転送の不足分があるならば，不足分の要求を作成し，転送命令を再発行する。

外部記憶装置に対するＣＰＵ命令としては，例えば以下のような命令が用意されている。

(a) 主記憶から外部記憶への仮想アドレスの指定によるデータ転送を指示する命令。

(b) 外部記憶から主記憶への仮想アドレスの指定によるデータ転送を指示する命令。

(c) 主記憶から外部記憶への実アドレスの指定によるデータ転送を指示する命令。

(d) 外部記憶から主記憶への実アドレスの指定によるデータ転送を指示する命令。

(e) 実際に転送された長さを求める命令。

このうち(a)～(d)は，ＣＰＵが要求の完了を待たな

い非同期命令であり，(e)はＣＰＵが要求の完了を待つ同期命令である。

例えば仮想計算機ＶＭ－Ａが，主記憶装置と外部記憶装置間で，各８Ｇバイトの転送要求を連続的に行い，仮想計算機ＶＭ－Ｂが，各４Ｍバイトの転送要求を連続的に行うものとする。

この場合，従来技術によるデータ転送のタイムチャートは，第６図（イ）に示すようになる。タイムチャートにおける実線部分は，データ転送中を示し，点線部分は，同一パスを使用するために転送待ちであることを示す。

従来技術の場合，各仮想計算機ＶＭ－Ａ，ＶＭ－Ｂからの転送要求を，そのままＶＭモニタが代行するため，仮想計算機ＶＭ－Ａの要求による８Ｇバイトのデータ転送が終了してから，仮想計算機ＶＭ－Ｂの４Ｍバイトのデータ転送要求が実行され，これが交互に繰り返される。したがって，特に仮想計算機ＶＭ－Ｂが，データベース処理やページング処理などの高速性を必要とする処理を行っている場合などには，転送待ち時間が長く，性能劣化の影響が大きかった。

これに対し，本発明の場合，仮想計算機ＶＭ－Ａに対してあらかじめ定義された転送優先度に応じて，８Ｇバイトのデータ転送要求を，ＶＭモニタにより，例えば５１２Ｍバイトの転送要求に分割して，転送命令を代行する。

したがって，第６図（ロ）に示すように，各４Ｍバイトの転送要求に対して，仮想計算機ＶＭ－Ｂが待たなければならない転送待ち時間は，５１２Ｍバイトの転送時間であり，転送待ち時間が短縮される。

〔発明の効果〕

以上説明したように，本発明によれば，仮想計算機上で動作するオペレーティング・システムからの主記憶装置と外部記憶装置間のデータ転送要求をそのまま実行しないで，仮想計算機モニタにより，１回当たりのデータ転送量を仮想計算機ごとに制御して実行する。したがって，仮想計算機ごとの性能に大きな偏りが生じることがなく，バランスのよいサービスが可能になる。

４．図面の簡単な説明

第１図は本発明の原理説明図，

第２図は本発明の第一実施例処理構成図，

第３図は本発明の第一実施例処理フロー，

第４図は本発明の第二実施例処理構成図，

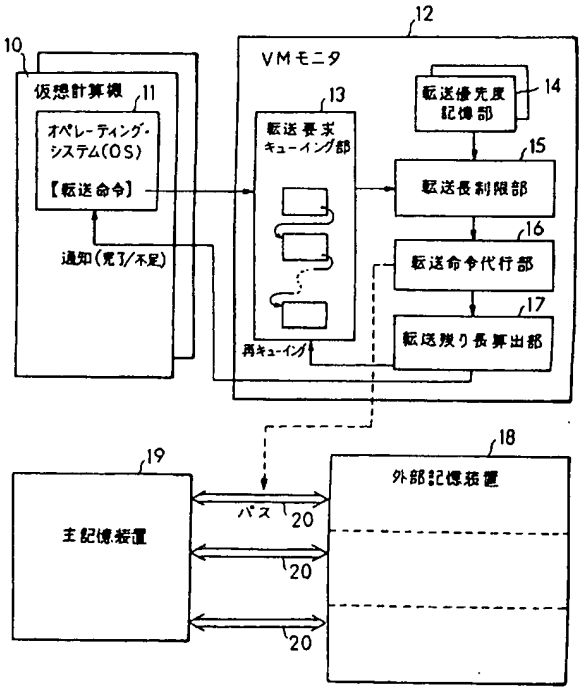
第５図は本発明の第二実施例処理フロー，

第６図は従来技術と本発明を比較するためのタイムチャートを示す。

図中，１０は仮想計算機，１１はオペレーティング・システム，１２はＶＭモニタ，１３は転送要求キューイング部，１４は転送優先度記憶部，１５は転送長制限部，１６は転送命令代行部，１７は転送残り長算出部，１８は外部記憶装置，１９は主記憶装置，２０はパスを表す。

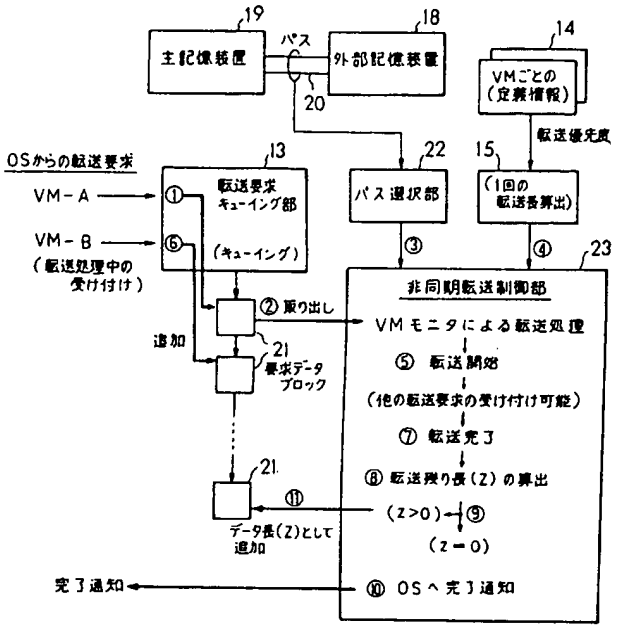
特許出願人　富士通株式会社

代理人　弁理士　小笠原吉義（外２名）



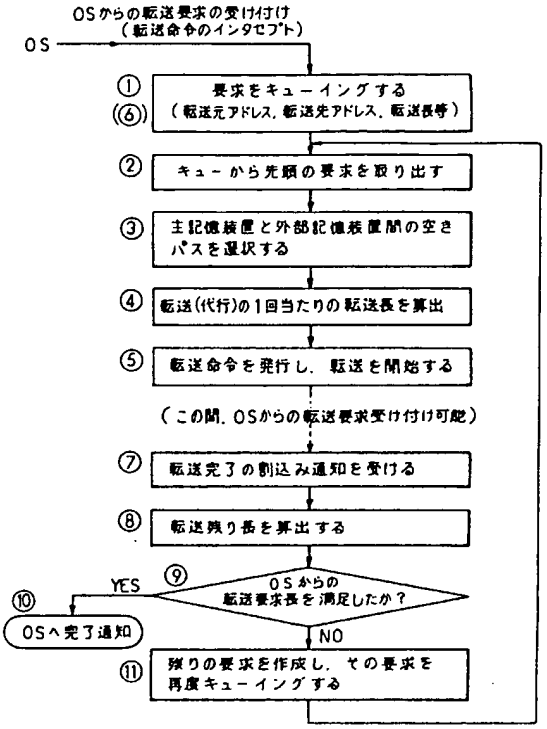
本発明の原理説明図

第　１　図



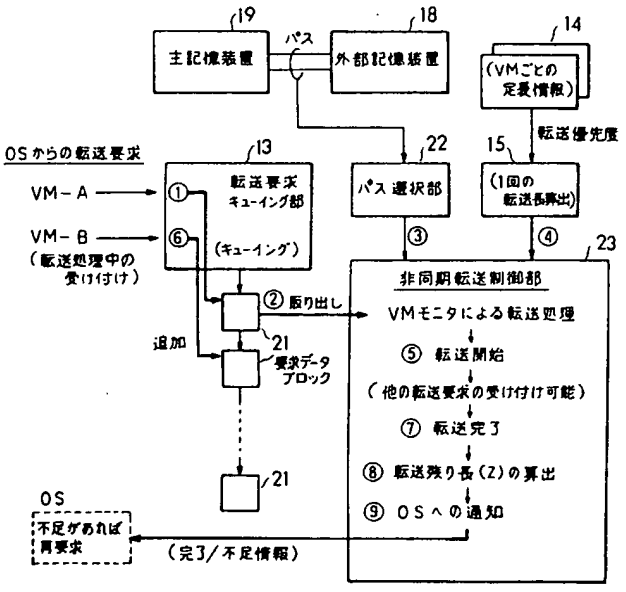
第一実施例処理構成図

第 2 図



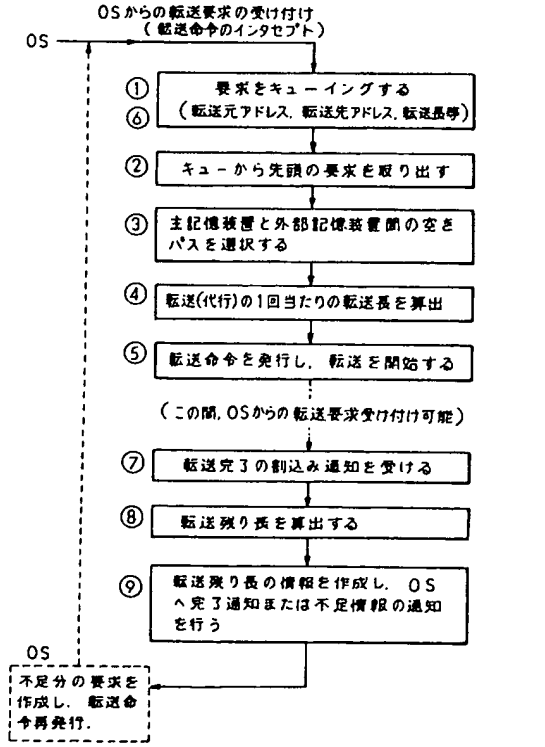
第一実施例処理フロー

第 3 図



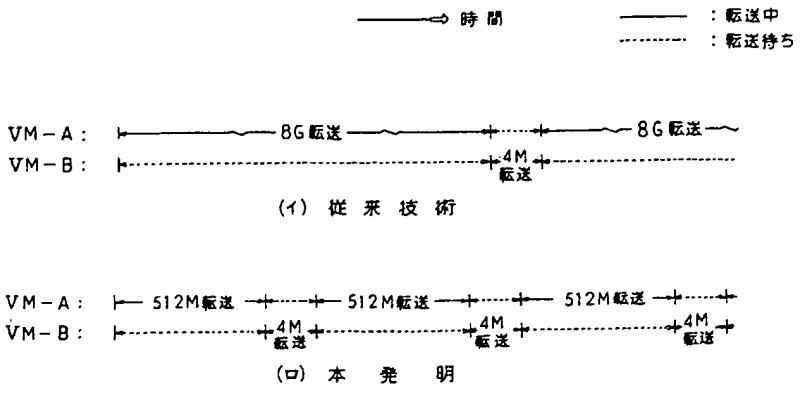
第二実施例処理構成図

第 4 図



第二実施例処理フロー

第 5 図


時間
：転送中
：転送待ち
VM－A：
8G転送
8G転送
VM－B：
4M
転送
(イ) 従 来 技 術
VM－A：
512M転送
512M転送
512M転送
VM－B：
4M
転送
4M
転送
4M
転送
(ロ) 本 発 明

タイムチャート

第 6 図